A PARTIAL LEAST SQUARES BASED RANKER FOR FAST AND ACCURATE AGE ESTIMATION

Hong Liu, Xiaohu Sun

Key Laboratory of Machine Perception,
Shenzhen Graduate School, Peking University, China
{hongliu, xiaohusun}@pku.edu.cn

ABSTRACT

Facial age estimation is challenging due to complex dynamics in aging process, which render metric regression methods unfavorable. Rankers show better performance by exploiting the ordinal nature of ages. The difficulty of designing a ranker is that each binary classifier of a ranker has to be trained using highly unbalanced positive and negative data. This paper proposes a partial least squares based ranker (PLS-Ranker), which fully maintains the advantages of PLS and greatly boosts its performance on the ordinal problem. In PLS-Ranker, an adaptive threshold learning strategy is proposed to boost each of the binary classifiers learned from highly unbalanced data. Previous ranking approaches such as CS-OHRank suffer from heavy computations because dozens of binary classifiers are trained separately. However, in PLS-Ranker, they are jointly learned. Additionally, PLS-Ranker simultaneously reduces feature dimensions and ranks in high speed even for high-dimensional features. Experimental results on the age estimation problem show that PLS-Ranker outperforms the state-of-the-art methods in terms of both accuracy and speed. PLS-Ranker also achieves state-of-the-art performance on the multi-source cross-race-and-gender age estimation problem, which further demonstrates its robustness.

***Index Terms*—** Age estimation, rank, partial least squares, adaptive threshold learning, speed

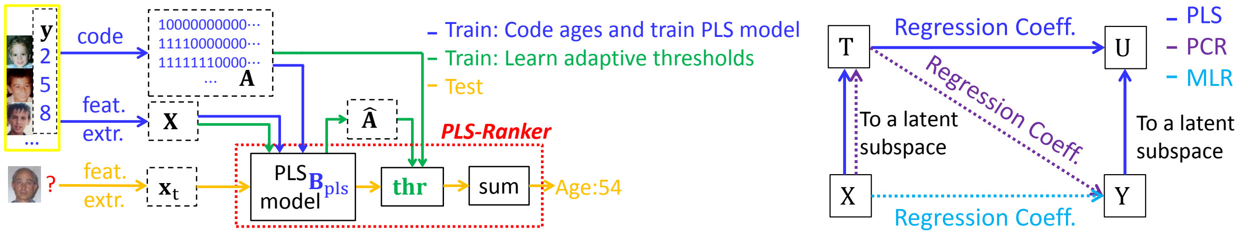
1. INTRODUCTION

Age estimation is useful for friendly and secure human-robot/computer interactions, age-based access control, family photo management, etc [1–3]. Multi-class classifiers take different age labels as totally independent. They often show poor performance due to ignoring the relative ordinal relation between labels [4]. Metric regression methods, e.g. warped gaussian process (WGP) and support vector regression (SVR), use real-valued labels [2, 5]. Metric regression methods also show unfavorable performance. The reasons may be two folds: 1) There exist complex dynamics influenced by many factors in aging. Although briefly speaking, aging in face is bone growth during childhood whereas skin related deformation during adulthood [4]; 2) Metric distance information carried by age is inexact [4, 5]. For instance, we can not say a 25-year-old adult is 2.5 times older than a 10-year-old child.

Relation to prior work. 1) Partial least squares. Although partial least squares (PLS) [6–8] has achieved success in chemometrics, it has been catching attention in computer vision area only in recent years [9–12]. The successful applications include human detection [9], face recognition [12] and pose estimation [11]. However, to some extent, PLS also suffers from the problems that the other metric regression methods encounter when being applied to the age estimation problem. **2) Ranking approaches.** Based on the fact that age has a nature order, Chang et al. [4, 13] proposed to use ranking methods for age estimation. Ranking approaches show advantages over metric regression methods such as WGP, SVR etc [4]. Rankers rank by ordinal comparisons (binary classifications) [5]. Among these ranking approaches [4, 13–15], cost-sensitive ordinal hyperplane rank (CS-OHRank) gets the minimum error [4]. The key to the accuracy of CS-OHRank is its basic binary classifier, cost-sensitive support vector machine (CS-SVM). CS-SVM is robust to *the imbalance problem of its positive and negative training data* (which will be described in detail in Sect. 2.2) using cost-sensitive strategy. However, CS-OHRank is time-consuming and of high complexity. A reason is that during the training process of CS-OHRank, dozens of rbf-kernel CS-SVMs have to be trained separately.

Since PLS has many excellent characteristics, our motivation is to transform the advanced multivariate data analysis (MVA) tool into a ranker to boost the performance in terms of accuracy, speed and robustness, on the age estimation prob-

This work is supported by National Natural Science Foundation of China (NSFC, No.61340046, 60875050, 60675025), National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247), Specialized Research Fund for the Doctoral Program of Higher Education (No.20130001110011), Natural Science Foundation of Guangdong Province (No.2015A030311034), Science and Technology Innovation Commission of Shenzhen Municipality (No.JCYJ20130331144631730, No.JCYJ20130331144716089).



(a)Flow chart of age estimation using PLS-Ranker (b) MLR, PCR and PLS coefficients calculation

Fig. 1. Partial least squares based ranker. **Best viewed in color.**

lem. We design a PLS based ranker (PLS-Ranker), in which all binary classifiers are jointly learned. We propose an adaptive threshold learning strategy to make each basic binary classifier of PLS-Ranker robust to the imbalance problem of its positive and negative training data.

2. PARTIAL LEAST SQUARES BASED RANKER

2.1. Partial least squares

A brief mathematical description of PLS is provided below. Please refer to [6–8,11,16] for additional details regarding advantages of PLS over multiple linear regression (MLR), principal components regression (PCR) and canonical correlation analysis (CCA).

The regressor (input) matrix $\mathbf{X}$ $(n \times N)$ and response (output) matrix $\mathbf{Y}$ $(n \times M)$ contain n observations of N independent variables and M dependent variables, respectively. PLS decomposes the zero-mean matrix $\mathbf{X}$ and zero-mean matrix $\mathbf{Y}$ into:

$$\begin{aligned} \mathbf{X} &= \mathbf{T}\mathbf{P}^T + \mathbf{E} \\ \mathbf{Y} &= \mathbf{U}\mathbf{Q}^T + \mathbf{F} \end{aligned} \tag{1}$$

where $\mathbf{T}$ and $\mathbf{U}$ are $n \times p$ matrices of the p extracted scores (a.k.a. components, latent variables or factors). $\mathbf{P}$ $(N \times p)$ and $\mathbf{Q}$ $(M \times p)$ are matrices of loadings. $\mathbf{E}$ and $\mathbf{F}$ are the residuals. The PLS method, whose classical form is based on the nonlinear iterative partial least squares (NIPALS) [6, 7], finds weights vector $\mathbf{w}$ and $\mathbf{c}$ such that:

$$\begin{aligned} [cov(\mathbf{t}, \mathbf{u})]^2 &= [cov(\mathbf{Xw}, \mathbf{Yc})]^2 \\ &= max_{|\mathbf{r}|=|\mathbf{s}|=1}[cov(\mathbf{Xr}, \mathbf{Ys})]^2 \end{aligned} \tag{2}$$

where $cov(\mathbf{t}, \mathbf{u}) = \mathbf{t}^T\mathbf{u}/n$ denotes the covariance between $\mathbf{t}$ and $\mathbf{u}$. Most PLS models assume that there exists a linear inner relation between the score vectors $\mathbf{t}$ and $\mathbf{u}$. Finally:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F}^* \tag{3}$$

where $\mathbf{B} = \mathbf{X}^T\mathbf{U}(\mathbf{T}^T\mathbf{X}\mathbf{X}^T\mathbf{U})^{-1}\mathbf{T}^T\mathbf{Y}$ is the regression coefficient matrix and $\mathbf{F}^*$ is the matrix of residuals [16]. For prediction, just multiply the feature block of test set $\mathbf{X}_{test}$ with $\mathbf{B}$.

The complexity of NIPALS is $O(n^2)$ [11], which is very fast in training. When using the kernel extension of NIPALS, kernel PLS (KPLS) [16], the complexity increases to $O(n^2N)$ [11], where N is the feature dimension. When N is big, which is often the case in applications, training KPLS model is expensive compared with training linear PLS model.

2.2. Partial least squares based ranker

The coding strategy for jointly learning. It is the first step of training the PLS-Ranker model, which is illustrated in Fig. 1(a) using blue lines. Given n training samples, the feature matrix $\mathbf{X}$ $(n \times N)$ and the column vector $\mathbf{y}$ $(n \times 1)$ contain feature vectors and scalar age values of the n samples. Suppose the age range is $1 \sim M_a$.

We encode each scalar age y_i $(i = 1, 2, \cdots, n)$ into a $1 \times (M_a - 1)$ binary row vector $\mathbf{A}(i, :)$ and form an $n \times (M_a - 1)$ indicator matrix $\mathbf{A}$:

$$A_{i,j} = \begin{cases} 1, & \text{if } j < y_i \\ 0, & \text{if } j \geqslant y_i \end{cases} \tag{4}$$

Elements in the j-th column of $\mathbf{A}$ indicate whether the n faces are older than j years or not. We propose the coding strategy to jointly learn all associated binary classifiers.

After the coding step, we use the feature matrix $\mathbf{X}$ and indicator matrix $\mathbf{A}$ to train a linear PLS model and get the regression coefficient matrix $\mathbf{B}_{pls}$. $\mathbf{B}_{pls}$ is the first parameter of the PLS-Ranker model.

The key to improve ranking accuracy. It is an effective and popular framework to reduce ranking to associated binary classifications [5]. Consider that we want to know how old a face is. An associated question would be:“ Is the face older than j?”. For a fixed j, such a question is exactly a binary classification problem, and the rank can be determined by asking multiple questions for $j = 1, 2, \cdots, (M_a - 1)$ [4,5].

Under the reduction framework, the key to improve ordinal ranking is to improve binary classifications [5]. However, the positive and negative training samples for each of the binary classifiers are highly unbalanced even though the age label (rank) distribution on the database is uniform. For example, suppose we have an aging database containing same number of faces for each age. Suppose the age range is 1∼77. For the 5-th binary classification problem: “Is the face older than 5?”, the number of positive (older than 5) and negative (not older than 5) training samples are of big difference.

Adaptive threshold learning. PLS-Ranker is also under the reduction framework. In PLS-Ranker, the trained linear PLS model will map the feature vector of a test sample to a *real-valued* indicator vector. Thus, sole linear PLS can not serve as the multiple binary classifiers without thresholding the real values to binary values.

The unbalanced positive and negative training samples may shift the optimal thresholds away from 0.5.[1] We propose an adaptive strategy to learn thresholds from the unbalanced training data for the $M_a - 1$ binary classifiers. It is illustrated in Fig. 1(a) using green lines. We apply the trained linear PLS model back to the feature matrix $\mathbf{X}$ of training set, and get a prediction of indicator matrix $\mathbf{A}$:

$$\hat{\mathbf{A}} = \mathbf{X}\mathbf{B}_{pls}. \tag{5}$$

where $\hat{\mathbf{A}}$ is a real-valued matrix. Thresholds are searched in a small range $\mathcal{S}$ around 0.5 and determined by the criterion of minimizing the error rate on the training set. For the j-th ($j = 1, 2, \cdots, M_a - 1$) binary classifier:

$$thr_j = \arg\max_{b \in \mathcal{S}} \sum_{i=1}^{n} f_m(A_{i,j}) \times f_m([\![\hat{A}_{i,j} \geqslant b]\!]) \tag{6}$$

where $[\![x]\!] = 1$ if x is true and $[\![x]\!] = 0$ otherwise. $f_m(x) = (x - 0.5) \times 2$, $x = 0, 1$ maps 0 and 1 to -1 and 1, respectively.[2] If the classification result is right for a sample, the summation will be increased by 1, otherwise decreased by 1. $\mathbf{thr}$ $(1 \times (M_a - 1))$ is the second parameter of PLS-Ranker.

Test process. The test process is illustrated in Fig. 1(a) using orange line. Feature vector of the test sample $\mathbf{x}_t$ $(1 \times N)$ is mapped to an indicator vector $\hat{\mathbf{a}}_t$ $(1 \times (M_a - 1))$ using the trained linear PLS model:

$$\hat{\mathbf{a}}_t = \mathbf{x}_t \mathbf{B}_{pls} \tag{7}$$

Now elements in $\hat{\mathbf{a}}_t$ are real-valued. We should threshold each element of $\hat{\mathbf{a}}_t$ into a binary indicator:

$$\hat{\mathbf{a}}_t(1, j) := [\![\hat{\mathbf{a}}_t(1, j) \geqslant thr_j]\!],\ j = 1, 2, \cdots, M_a - 1 \tag{8}$$

The rank (age) is obtained by summarizing elements in $\hat{\mathbf{a}}_t$:

$$\hat{r}_t = \sum_{j=1}^{M_a - 1} \hat{\mathbf{a}}_t(1, j) + 1 \tag{9}$$

Time cost. Considering the time cost, linear PLS rather than kernel PLS is used. Specially, the linear PLS algorithm, SIMPLS [6, 8], rather than the classical NIPALS, is adopted in our experiments. SIMPLS has several advantages over NIPALS and one of them is not involving a breakdown of the $\mathbf{X}$ matrix, and for this reason it is faster [6, 8].

Simultaneous dimensionality reduction and ranking. In the PLS-Ranker, the score matrix $\mathbf{T}$ $(n \times p)$ is actually the dimensionality reduction matrix of $\mathbf{X}$ $(n \times N)$. Usually $p \ll N$ holds, and more latent variables (big p) are not necessarily in practice (see Fig. 2 and Table 1, 4), thus PLS-Ranker is very fast in practice.

3. EXPERIMENTS AND DISCUSSIONS

3.1. Datasets and settings

FG-NET[3] contains 1002 images of 82 persons. Leave-one-person-out (LOPO) setting is used [4, 17].

[1] 0.5 is the intermediate value of 0 and 1. Empirically, we can chose 0.5 as the threshold when numbers of the positive and negative samples used to train the binary classifier are equal.

[2] If not mapping 0 to -1, there will be no difference between right classification $0 \times 0 = 0$ and wrong classification $1 \times 0 = 0$.

[3] http://www-prima.inrialpes.fr/FGnet/

MORPH. There exist different test settings for MORPH [18]. Both of the following settings are used in our experiments to fully compare with the state-of-the-art methods.

Setting 1 selects 55132 images from the original MORPH and categorizes them into S1 (10530), S2 (10530) and S3 (34072) [19, 20]. Firstly S1 is used for training and S2+S3 for testing, then S2 for training and S1+S3 for testing [10].

Setting 2 selects 5,493 images of Caucasian (Cau)[4] from the original MORPH [4]. 80% images of Cau are randomly selected as training data and the rest as testing data [3, 4].

Note that each of S1 and S2 contains 3980 Black Males (BM), 3980 White Males (WM), 1285 Black Females (BF) and 1285 White Females (WF). We use S1 and S2 to study cross-race-and-gender age estimation. The percentage of training data in the target population is 50% [20]. Please refer to [19, 20] for further details about the test strategy.

Evaluation metric is mean absolute error (MAE), i.e. the average absolute error between predicted age value and ground truth over all testing samples.

Features. Bio-inspired features (BIF) [1] are extracted from MORPH (Setting 1) (4376D) and Cau (4376D)[5]. Active Appearance Model features (AAM) [21] are extracted from FG-NET (492D) and Cau (886D). AAM inherently uses PCA. Large percent of variances are preserved when extracting the AAM features because AAM is mainly used for extracting raw low-level features rather than for dimensionality reduction, on which task our PLS-Ranker performs better.

3.2. Results and discussions

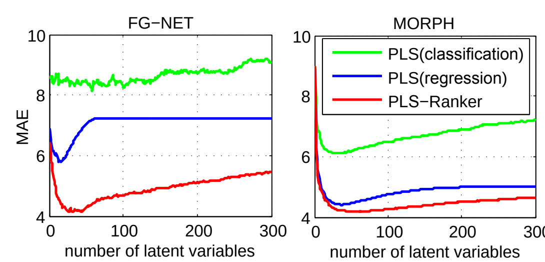


Fig. 2. PLS-Ranker vs. PLS

Table 1. PLS-Ranker vs. PLS

Method	MAE/year (num. of latent var.)	
	FGNET	MORPH
PLS (classification)	8.14 (58)	6.10 (33)
PLS (regression)	5.78 (17)	4.40 (37)
PLS-Ranker	**4.14** (45)	**4.17** (49)

PLS-Ranker vs. PLS. PLS can be used for classification and metric regression. The usage for classification is inserting a 1 into a row vector $\mathbf{0}$ $(1 \times M_a)$ to indicate membership, and taking the vector as the output [6]. We compare PLS-Ranker with the two usages of PLS on FG-NET and MORPH (Setting 1). Using different number of latent variables, PLS-Ranker consistently outperforms PLS (see Fig. 2). Minimums of the

[4] http://www.iis.sinica.edu.tw/~kuangyu/OHRank.htm

[5] We use BIF provided by [19, 20]. Intersection of the MORPH (Setting 1) [19, 20] and Cau [4] contains 5444 images.

curves are tabulated in Table 1. PLS-Ranker performs well on both big and small databases. However, the performance of PLS for classification or regression degrades severely.

Guo et al. [10] achieves state-of-the-art result MAE=4.18 using kernel PLS complying with Setting 1 on MORPH. However, kernel PLS has high computational cost compared with linear PLS, as emphasized in Sect. 2.1. For this reason, they firstly use linear PLS for feature dimension reduction and then employ kernel PLS for regression. However, the performance is degraded to MAE=4.43 [10].

Note that although linear PLS is adopted in PLS-Ranker for high speed, it is interesting and easy to substitute linear PLS with kernel PLS, and even CCA, kernel CCA or other MVA approaches to test our transform-to-ranker strategy. However, it is not the focus of this paper.

Results on FG-NET and MORPH. We compare PLS-Ranker with the state-of-the-art methods on FG-NET and MORPH (Setting 2) in terms of both speed and accuracy. CA-SVR [15] uses similar strategy to encode the numeric age. However, when mapping feature vector to the attribute-space, it uses similar method as MLR (Fig. 1(b)), which is a naive MVA approach. Furthermore, after mapping, another regression model has to be trained to map the point in attribute-space to a scalar value [15]. CS-OHRank, although gets lower MAE than CA-SVR, is time-consuming. PLS-Ranker outperforms the complex feature combination method [22] using only AAM features. PLS-Ranker also outperforms the deep learning based method proposed recently [3]. In [3], SVR is used for regression after extracting raw features using CNN, which may harms its performance.

PLS-Ranker can simultaneously reduce feature dimension and rank in high speed even for high-dimensional features like BIF. The number of latent variables (components), i.e. the feature dimension after reduction is often less than 100. The training time is the minimum compared with other methods in the literatures [4, 15], two orders of magnitude (10^2) faster than CA-SVR. The experimental results on FG-NET and MORPH demonstrate that PLS-Ranker outperforms the state-of-the-art methods in terms of both speed and accuracy.

Cross-race-and-gender age estimation. People from different races, with different genders, may age differently. Only a few works focus on cross-race and cross-gender age estimation (Crg) [19, 20]. In this paper, we study the multi-source cross-race-and-gender (ms-Crg) age estimation problem. In the problem, training set mainly contains faces of other races or gender, and only a smaller number of samples with the same race and gender as the testing faces. As CpDA [20] achieves state-of-the-art results on the problem, we compare our method with it. Note that our PLS-Ranker does not exploit the race information as CpDA. The MAE is reduced 25.31% from 5.99 to 4.48 years, which outperforms the state-of-the-art by a large margin. It demonstrates that PLS-Ranker is robust to race and gender variations and can extract latent stable and significant aging features.

Table 2. Comparison with the state-of-the-art methods

Method	MAE/year	
	FGNET	MORPH
MTWGP [2]	4.83	6.28
PLO [14]	4.82	–
AAM+CA-SVR [15]	4.67	5.88
Feat. combine + select [22]	4.49	–
AAM+CS-OHRank [4]	4.48	6.07
Regularized CA-SVR [17]	4.37	–
Deep Feature+SVR [3]	4.26	4.77
AAM+PLS-Ranker	**4.14**	**5.38**
BIF+PLS-Ranker	–	**3.77**

Table 3. Training time required by different models. Tested on an Intel(R) Core i5-3470 (3.2GHz), 8G RAM PC.

Method	Training time/min	
	FGNET	MORPH
OHRank [4]	1.30×10^4	3.02×10^4
SVR [23]	2.69×10^0	2.08×10^1
CA-SVR [15]	8.91×10^{-1}	6.10×10^0
AAM+PLS-Ranker	$\mathbf{7.20 \times 10^{-3}}$(0.43s)	$\mathbf{2.25 \times 10^{-2}}$(1.35s)
BIF+PLS-Ranker	–	$\mathbf{1.25 \times 10^{-1}}$(7.51s)

Table 4. Results on ms-Crg age estimation problem

Train	Test	MAE/year (num. of latent var.)	
		CpDA [20]	PLS-Ranker
BF+WF	BM	6.47	**4.55** (27)
BF+WF	WM	5.70	**3.87** (49)
WM+BM	WF	6.58	**5.10** (80)
WM+BM	BF	6.40	**5.49** (67)
BF+BM	WF	6.59	**5.24** (88)
BF+BM	WM	5.23	**3.85** (63)
WF+WM	BF	6.32	**5.65** (39)
WF+WM	BM	5.96	**4.49** (31)
Average		5.99	**4.48 (25.31%)**

4. CONCLUSIONS

This paper proposes a partial least squares based ranker, PLS-Ranker. The adaptive threshold learning strategy makes each binary classifier of PLS-Ranker robust to the imbalance problem of its training data, thus boosts ranking accuracy of PLS-Ranker. Dozens of binary classifiers are jointly learned through the proposed coding strategy. PLS-Ranker simultaneously reduces feature dimension and ranks in high speed even for high-dimensional features, which makes it suitable for large-scale ordinal ranking problems and real-time applications. It is easy to implement PLS-Ranker using existing PLS tools. Experimental results show that PLS-Ranker outperforms the state-of-the-art methods for age estimation in terms of both accuracy and speed. PLS-Ranker also achieves state-of-the-art result on the ms-Crg age estimation problem, which further demonstrates its robustness. In future work, we plan to apply the coding and adaptive threshold learning strategy to other MVA methods such as KPLS, CCA, KCCA, and try to boost their performances on ranking problems.

5. REFERENCES

[1] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *CVPR*, 2009, pp. 112–119.

[2] Y. Zhang and D. Yeung, "Multi-task warped gaussian process for personalized age estimation," in *CVPR*, 2010, pp. 2622–2629.

[3] X. Wang, R. Guo, and C. Kambhamettu, "Deeply-learned feature for age estimation," in *WACV*, 2015, pp. 534–541.

[4] K. Chang, C. Chen, and Y. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *CVPR*, 2011, pp. 585–592.

[5] H. T. Lin, *From ordinal ranking to binary classification*, Ph.D. thesis, California Institute of Technology, 2008.

[6] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Subspace, latent structure and feature selection*, pp. 34–51. Springer, 2006.

[7] H. Wold, "Soft modeling by latent variables: the non-linear iterative partial least squares approach," *Perspectives in probability and statistics, papers in honour of MS Bartlett*, pp. 520–540, 1975.

[8] S. De Jong, "SIMPLS: an alternative approach to partial least squares regression," *Chemometrics and intelligent laboratory systems*, vol. 18, no. 3, pp. 251–263, 1993.

[9] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *ICCV*, 2009, pp. 24–31.

[10] G. Guo and G. Mu, "Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression," in *CVPR*, 2011, pp. 657–664.

[11] M. A. Haj, J. Gonzalez, and L. S. Davis, "On partial least squares in head pose estimation: How to simultaneously deal with misalignment," in *CVPR*, 2012, pp. 2602–2609.

[12] A. Sharma and D. W. Jacobs, "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch," in *CVPR*, 2011, pp. 593–600.

[13] K. Chang, C. Chen, and Y. Hung, "A ranking approach for human ages estimation based on face images," in *ICPR*, 2010, pp. 3396–3399.

[14] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," in *CVPR*, 2012, pp. 2570–2577.

[15] K. Chen, S. Gong, T. Xiang, and C. C. Loy, "Cumulative attribute space for age and crowd density estimation," in *CVPR*, 2013, pp. 2467–2474.

[16] R. Rosipal and L. J. Trejo, "Kernel partial least squares regression in reproducing kernel hilbert space," *The Journal of Machine Learning Research*, vol. 2, pp. 97–123, 2002.

[17] Q. Tian and S. Chen, "Cumulative attribute relation regularization learning for human age estimation," *Neurocomputing*, 2015.

[18] K. Ricanek Jr and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *FGR*, 2006, pp. 341–345.

[19] G. Guo and G. Mu, "Human age estimation: What is the influence across race and gender?," in *CVPRW*, 2010, pp. 71–78.

[20] G. Guo and C. Zhang, "A study on cross-population age estimation," in *CVPR*, 2014, pp. 4257–4263.

[21] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models," *TPAMI*, no. 6, pp. 681–685, 2001.

[22] H. Ren and Z. N. Li, "Age estimation based on complexity-aware features," in *Computer Vision–ACCV 2014*, pp. 115–128. Springer, 2015.

[23] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *TIP*, vol. 17, no. 7, pp. 1178–1188, 2008.